
دار المنظومة
DAR ALMANDUMAH
الرواد في قواعد المعلومات العربية

الدلالة الاكلينيكية للبحوث النفسية والتربوية المستخدمة للتصاميم التجريبية "دراسة تحليلية"

يحيى حياتي نصار*

جامعة الإمارات العربية المتحدة، الإمارات العربية المتحدة

استلم بتاريخ: ٢٠١٦/١٠/٣١ قبل بتاريخ: ٢٠١٦/١١/٢٣

ملخص: هدفت الدراسة الحالية بصورة أساسية إلى إبراز أهمية مفهوم الدلالة الإكلينيكية والتمييز بينه وبين الدلالة الإحصائية والدلالة العملية لنتائج البحوث النفسية والتربوية وخاصة في حالة الدراسات المستخدمة للتصاميم التجريبية. كما هدفت الدراسة الحالية إلى تحديد أهم الطرق التي يمكن أن تستخدم لتقدير الدلالة الإكلينيكية، حيث تم استعراض طريقة جاكبسون- ترواكس Jacobson-Truax ((JT)، وطريقة جوليكسن- لورد- نوفك Gulliksen-Lord-Novick (GLN)، وطريقة أدواردز- نانولي Edwards-Nunnally ((EN)، وطريقة هيجمان- أريندل Hageman-Arrindell (HA)، وأخيرا الطريقة الخطية الهرمية Hierarchical Linear Method (HLM). حيث تم في هذه الدراسة التركيز على طريقة جاكبسون- تراوكس (JT) ومن خلال استخدام بيانات افتراضية تم تقديم مثالين: نفسي وتربوي تم من خلالهما توضيح كيفية تطبيق هذه الطريقة. كما حاولت الدراسة الحالية من خلال توظيف طريقة جاكبسون- تراوكس (JT) تقديم نموذج مقترح لفحص الدلالة الإكلينيكية لأثر المعالجة التجريبية على مستوى الفروق بين المجموعات التجريبية والضابطة. أشارت نتائج الدراسة لأهمية فحص الدلالة الإكلينيكية للبحوث النفسية والتربوية حيث يمكن أن تكون النتائج دالة إحصائيا وعمليا ولكنها غير دالة من الناحية الإكلينيكية سواء على مستوى الأفراد أو على مستوى المجموعات.

كلمات مفتاحية: الدلالة الإكلينيكية، الدلالة الإحصائية، الدلالة العملية، طريقة جاكبسون_تراوكس.

Clinical Significance of Psychological and Educational Experimental Studies "Analytical Study"

Yahya H. Nassar*
United Arab Emirates University, United Arab Emirates

Abstract: The main aim of this study was to confirm the importance of the concept of clinical significance. Also this study aimed at discriminating among the concepts of statistical, practical and clinical significance. In addition, this study attempted to present the most known methods to estimate the clinical significance of psychological and educational studies. These methods were Jacobson-Truax method (JT), GulliksenLord-Novick method (GLN), Edwards-Nunnally method (EN), Hageman-Arrindell method (HA), and Hierarchical Linear Method (HLM). The current study constraints on the Jacobson-Truax method (JT) and via using hypothetical data, psychological and educational examples were presented to explain how to implement JT method. Moreover, via applying JT method this study attempted to introduce a suggested model to investigate the clinical significance of the treatment on the differences among the groups of the experimental designs. Finally, the results of this study confirmed the value of investigating the clinical significance of the psychological and educational studies because the results could be statistically and practically significant although could not indicate to clinical significance either on the individual or group level.

Keywords: Clinical significance, statistical significance, practical significance, Jacobson-Truax method.

*yanassar@yahoo.com

يهتم الباحثون في مجال الدراسات النفسية والتربوية بصورة أساسية باستخراج الدلالة الإحصائية لنتائج أبحاثهم، حيث أن هناك فكرة سائدة بين الباحثين في مجال الدراسات النفسية والتربوية أن النتائج الجيدة هي النتائج التي تحقق الدلالة الإحصائية وعلى العكس من ذلك فإن النتائج تعتبر ليست ذات قيمة إذا كانت غير دالة إحصائيا. إن مفهوم الدلالة الإحصائية يشير إلى أن دور عامل الصدفة في الفروق بين متوسطات المجموعات التجريبية والضابطة أقل من الحد المسموح به من قبل الباحث (Tuckman, 2005). وتشير الدراسات إلى أن مفهوم أو إستراتيجية الدلالة الإحصائية قد تكون مضللة في كثير من الاحيان، حيث يمكن أن تكون النتائج دالة إحصائيا نتيجة لتكبير أو تضخيم حجم العينة، بل إنه يمكن تحويل البيانات غير الدالة إحصائيا الى دالة إحصائيا عن طريق زيادة حجم العينة (نصار، ٢٠٠٦). كما أن نتائج بعض الدراسات قد تكون دالة من الناحية الإحصائية لكنها غير دالة من الناحية العملية (الصياد، ١٩٨٩، النجار، ١٩٩٠) من هنا بدأ الحديث عن مفهوم آخر في الأبحاث النفسية والتربوية ألا وهو الدلالة العملية للنتائج أو ما يسمى بحجم الأثر (Effect size) (عودة والخليلي، ١٩٨٨).

ويشير مفهوم حجم الأثر بصورة أساسية إلى أن النتائج الدالة إحصائيا ليس بالضرورة أن تكون دالة من الناحية العملية، حيث يتم فحص الدلالة العملية لنتائج الأبحاث النفسية والتربوية من خلال أساليب إحصائية معينة مثل مؤشر "د" في حالة التعامل مع عينتين مستقلتين أو اختبار مربع ايتا إذا تضمنت الدراسة التجريبية متوسطين أو أكثر، أي أن هناك أسلوبا إحصائيا أو أكثر يمكن استخدامها لفحص الدلالة العملية للنتائج في البحوث النفسية والتربوية بحيث تكون مكملة لتلك النتائج التي يتم الحصول عليها من الاختبارات الإحصائية التي تفحص الدلالة الإحصائية للبيانات في الدراسات النفسية والتربوية (نصار، ٢٠٠٦). إن الموضوع الأساس والذي أغفلته الكثير من الدراسات النفسية والتربوية في الدول العربية، وربما على مستوى الأدب النفسي والتربوي بصورة عامة هو الدلالة الإكلينيكية لنتائج تلك الأبحاث. يشير مفهوم الدلالة الإكلينيكية للنتائج وخاصة تلك المستخدمة للتصاميم التجريبية إلى وجود فروق ليس فقط دالة إحصائيا أو/ ودالة عمليا بين المجموعات التجريبية والضابطة بل إلى أن البرامج المستخدمة في تلك الدراسات فعالة من حيث تغيير واقع حال أفراد المجموعة أو المجموعات التجريبية من حالة إلى حال أخرى من المتوقع أن تكون هي الأفضل أو المرغوب بها (Campbell, 2005). فعلى سبيل المثال إن إجراء دراسة تجريبية تهدف بصورة أساسية إلى تخفيض مستوى قلق الاختبار لدى طالبات الثانوية العامة في الزرقاء من خلال برنامج إرشاد جمعي نفسي يتم تطبيقه على أفراد المجموعة التجريبية بينما لا تتعرض المجموعة الضابطة لذلك البرنامج، هو مثال على الدراسات المستخدمة للتصاميم التجريبية في مجال الإرشاد النفسي والتربوي، حيث قد يقوم الباحث في مثل هذه الدراسة بفحص الدلالة الإحصائية للفرق بين متوسطي المجموعتين التجريبية والضابطة وفي أحسن الأحوال قد يقوم الباحث بفحص الدلالة العملية للنتائج في هذه الدراسة، لكن السؤال المطروح هنا هل كان البرنامج الإرشادي فعالا بحيث انخفض مستوى قلق الاختبار لدى الطالبات اللواتي تعرضن للبرنامج وبحيث تحولوا من طالبات يعانين من قلق الاختبار إلى طالبات عاديات يواجهن الاختبارات التحصيلية بكل ثقة وبدون أية أعراض غير عادية للقلق المعيق للأداء. وفي رأيي أن هذا هو الهدف الأساس للدراسات القائمة على برامج إرشادية أو تدريبية. إن الدلالة الإكلينيكية للنتائج وفقا للمفهوم المشار إليه تعتبر هامة لفهم وتفسير نتائج الأبحاث النفسية والتربوية، بل أنها قد تعتبر معيارا للحكم على التقدم أو التحسن أو التراجع الذي حصل لأفراد عينة الدراسة بعد تعرضهم للمعالجة (Lambert et al., 2001). كما أن أهمية مفهوم الدلالة الإكلينيكية تبرز من خلال اهتمامها بالتغيير

الذي حصل على مستوى كل فرد من أفراد الدراسة على حدة وبالتالي تحديد مدى استفادة كل فرد من هؤلاء الأفراد من المعالجة المستخدمة في الدراسة. وبالتالي فإن الدلالة الإكلينيكية تهتم باستكشاف أو اختبار الفائدة التي يمكن تحقيقها من خلال المعالجات المستخدمة في الدراسات التجريبية بطرق أكثر فائدة من الناحية العملية من الطرق التقليدية التي تعتمد على الدلالة الإحصائية والدلالة العملية (Hansen, Lambert, & Forman, 2002).

مشكلة الدراسة

إن الهدف الأساس من الدراسة الحالية هو تقديم مفهوم الدلالة الإكلينيكية للنتائج في الدراسات النفسية والتربوية المستخدمة للتصاميم التجريبية والتمييز بينه وبين مفاهيم الدلالة الإحصائية والدلالة العملية من ثم استعراض لأهم الأساليب المستخدمة في تقدير أو حساب الدلالة الإكلينيكية لنتائج الأبحاث التجريبية.

أسئلة الدراسة: حاولت الدراسة الحالية الإجابة على الأسئلة الآتية:

١. ما هو مفهوم الدلالة الإكلينيكية لنتائج البحوث النفسية والتربوية والمستخدمة للتصاميم التجريبية؟

٢. ما هو الفرق بين الدلالة الإحصائية والدلالة الإكلينيكية؟

٣. هل الدلالة الإكلينيكية هي نفسها الدلالة العملية للنتائج أم أن هناك فرق بينهما؟

٤. ما هي الأساليب أو الإجراءات والتي يمكن من خلالها التحقق من الدلالة الإكلينيكية لنتائج البحوث النفسية والتربوية والمستخدمة للتصاميم التجريبية؟

٥. هل يمكن استخدام الدلالة الإحصائية لفحص الدلالة الإكلينيكية للفروق بين متوسطات المجموعات التجريبية والضابطة؟

أهمية الدراسة

إن اعتبار الهدف الأساس من أي تحليل إحصائي هو اكتشاف فيما اذا كان هناك دلاله إحصائيه للنتائج يعتبر مشكله في حد ذاتها، لأنه على الرغم من أهمية الحصول على نتائج دالة إحصائيا، فإنها قد لاتكون كافية (Thompson, 2006) وكما أشار بعض الباحثين فإن مفاهيم "الدلاله الإحصائيه" و"الدلاله العملية" قد تداخلت واستخدمت في كثير من الأحيان بطريقه غير صحيحة (Hubbar and Ryan, 2000; Huberty , 2002). وفي هذا السياق، تشير دراسة الثبيتي (٢٠٠٨) إلى أن معظم الدراسات التجريبية تركز على مستوى الدلالة لاتخاذ قرار حول قبول الفرضية الصفرية أو رفضها دون الاهتمام باستخدام الدلالة العملية للنتائج. وقد ظهر حديثا في الدراسات الإنسانية أسلوب ثالث لتفسير نتائج الأبحاث وهو يعتمد بصوره أساسيه على ما يسمى بالدلالة الإكلينيكية (Campell, 2005)وقد استخدم هذا المفهوم لفحص فيما اذا كانت النتائج التي يتم الحصول عليها في الدراسات التجريبية تعتبر هامة فعلا وذات قيمة بالنسبة للموضوع او للظاهرة قيد الدراسة وبصورة اساسية تتركز أهمية الدراسة الحالية بأنها تحاول التمييز بين ثلاثة أنواع من الدلالات التي قد تستخدم في الدراسات النفسية و التربوية وخاصة في سياق التصاميم التجريبية وهذه الدلالات هي: الدلالة الاحصائية والدلالة العلمية والدلالة الإكلينيكية. كما أن الدراسة الحالية تسعى لتقديم أساليب إحصائية محددة لفحص الدلالة الإكلينيكية على مستوى الأفراد وهو أساس تميز هذا الأسلوب عن الدلالات الأخرى، كما حاولت الدراسة الحالية تقديم أو اقتراح نموذج أو أسلوب يتم من خلاله استخدام الدلالة الإحصائية لفحص الدلالة الإكلينيكية بين المجموعات التجريبية والضابطة. ولعل الأهمية الأساسية للدراسة الحالية تكمن في تقديمها طريقة مهمة للباحثين العرب لتفسير نتائج أبحاثهم التجريبية في مجال الدراسات الإنسانية وبالتحديد في مجال الدراسات النفسية والتربوية ألا وهي الدلالة

الإكلينيكية. ومن الجدير ذكره في هذا السياق أن هذا المفهوم قد ظهر عام ١٩٨٤ وأنه قد أصبح من العناصر الأساسية التي تتوقع بعض المجلات المتخصصة في مجال الدراسات النفسية والتربوية أخذها بعين الإعتبار في تفسير نتائج الأبحاث التجريبية والتي يكون الهدف الاساس أو الغرض منها العلاج أو التحسين (Kendall, 1999; Kendall, Marrs-Garcia, Nath, & Sheldrik, 1999)

تعريف مفاهيم الدراسة

الدلالة الإكلينيكية (Clinical Significance): مقدار التغير في سلوك الفرد أو المجموعة والناتج عن أثر المعالجة، حيث تكون النتائج دالة إكلينيكيا عند انتقال الفرد أو المجموعة من حالة إلى حالة أخرى، ويتم تقديرها سواء على مستوى الفرد أو المجموعة من خلال أساليب إحصائية سيتم استعراض بعضها في سياق الدراسة الحالية.

الدلالة الإحصائية (Statistical Significance): أسلوب إحصائي يستخدم لفحص دور أو مقدار احتمال عامل الصدفة في العلاقة بين المتغيرات أو في الفروق بين المتوسطات، حيث تكون النتائج دالة إحصائيا إذا كان الاحتمال الفعلي للصدفة أقل أو يساوي أعلى حد مسموح به لعامل الصدفة في العلاقة والذي يتم تقريره من قبل الباحث والمتضمن في الفرضيات الصفرية والبديلة أو يسمى بمستوى الدلالة المتوقع.

الدلالة العملية المتوقعة (Expected Practical Significance) أو حجم الأثر المتوقع: مقدار قوة العلاقة المتوقعة بين متغيرات الدراسة والتي يمكن تقديرها من خلال مراجعة الأدب السابق.

الدلالة العملية الفعلية (Real Practical significance) أو حجم الأثر الفعلي: مقدار قوة العلاقة الفعلية بين متغيرات الدراسة والتي يمكن تقديرها من خلال عملية تحليل بيانات الدراسة وباستخدام أساليب إحصائية محددة مثل كوهين "د" ومربع إيتا ($\eta2$) وغيرها (نصار، ٢٠٠٦).

الدراسة التحليلية: هي دراسة تقوم على تحليل ظاهرة نفسية أو اجتماعية أو طبيعية، أو مفهوم علمي سواء من حيث تعريفه وتوضيح علاقته بالمفاهيم المرتبطة به بطرق مختلفة والتي منها تقديم أمثلة علمية يتم فيها استخدام وتحليل بيانات افتراضية وهذه هي الطريقة التي استخدمت في الدراسة الحالية.

الطريقة والإجراءات

لتحقيق أهداف الدراسة الحالية وللإجابة على أسئلتها البحثية فإنه تم استخدام المنهج الوصفي التحليلي حيث تم استعراض أو توضيح مفهوم الدلالة الإكلينيكية من خلال التمييز بين هذا المفهوم من جهة ومفهومي الدلالة الإحصائية والعملية من جهة ثانية. كذلك فإنه تم في الدراسة الحالية محاولة تقديم بعض الأساليب الإحصائية الأكثر استخداما في الأدب النفسي والتربوي لاستكشاف الدلالة الإكلينيكية للأبحاث المستخدمة للتصاميم التجريبية. كذلك سيتم في هذه الدراسة تقديم أمثلة بحثية نفسية وتربوية افتراضية بغرض توضيح مفهوم الدلالة الإكلينيكية وتبيان كيفية تقديرها أو حسابها وتفسيرها خاصة للباحثين المهتمين باستخدام طرق غير تقليدية لتفسير نتائج أبحاثهم التجريبية. كذلك سيتم في الدراسة تقديم نموذج مقترح يهدف إلى فحص الدلالة الإكلينيكية للمعالجات التجريبية على مستوى المجموعات أي تحديد الدلالة الإكلينيكية لتلك المعالجات على المجموعات التجريبية والضابطة ، حيث يعتمد هذا النموذج على استخدام الدلالة الإحصائية ولكن بعد تصنيف أفراد المجموعتين التجريبية والضابطة وفقا للمعايير المستخدمة في واحدة من أشهر الطرق المستخدمة في تقدير الدلالة الإكلينيكية وهي طريقة جاكبسون- تروكس (Jacobson & Truax, 1991) ومن ثم استخدام مربع كاي لجداول التوافق (Chi- Square for tabulation cross) لفحص فيما إذا كان هناك علاقة دالة إحصائيا بين متغيري المجموعة (تجريبية – ضابطة) وتصنيف أفراد العينة تبعا للطريقة المذكورة، وسيتم تقديم مثال

توضيحي باستخدام بيانات افتراضية لتبيان أسلوب تطبيق النموذج المقترح.

الدلالة الإحصائية (Statistical Significance)

لقد ظهر مفهوم الدلالة الإحصائية منذ ثلاثمائة عام تقريبا (Thomopson, 2002) ولكن تم استخدامه في بدايات القرن العشرين بصورة اكثر كثافة، وخاصة عند استخدام الاختبارات الإحصائية مثل مربع كاي (Chi_ Sauare) واختبار "ت"، واختبار تحليل التباين الاحادي (ANOVA). ولقد انتشر استخدام أسلوب فحص الدلالة الإحصائية للفرضيات الصفرية في الآونة الأخيرة بصورة متزايدة، وكان الهدف من ذلك هو اتخاذ قرار فيما إذا كانت النتائج التي تم التوصل إليها دالة إحصائيا أم غير دالة.

ويقوم هذا الأسلوب على صياغة فرضية صفرية تبنى حول مجتمع الدراسة ويتم اختبارها من خلال بيانات عينة من المفترض انها تتشابه في خصائصها مع خصائص ذلك المجتمع. ويستخدم مفهوم مستوى الدلالة الفعلي لاختبار فيما اذا كانت الفروق بين متوسطات المجموعات التجريبية والضابطة دالة إحصائيا ام لا (Thomopson, 1998). إن المشكلة في استخدام الدلالة الاحصائية تكمن في أنها لا تشير إلى الأهمية الفعلية للنتائج، وبالتالي لا يستطيع الباحثون تحديد درجه أهميه نتائجهم في سياق موضوع الدراسة أو تبعا للظاهرة قيد الدراسة إنها فقط تخبر فيما إذا كانت النتائج التي تم الحصول عليها تعود إلى عامل الصدفة أم لا. وفي ضوء ذلك يستطيع الباحث أن يتعامل مع الفرضية البديلة التي تمثل وجهة نظره ، كما أنه في بعض الأحيان قد لا يكون قادرا على التعامل مع هذه الفرضية. ومن المعروف إحصائيا أن هذا النوع من الدلالة يتأثر بصورة مباشرة بحجم العينة، بمعنى أن النتائج غير الدالة إحصائيا يمكن أن تصبح دالة إحصائيا من خلال زيادة حجم العينة دون حدوث أي تغيير في خصائص التوزيع الإحصائي للبيانات الأصلية (قبل زيادة حجم العينة) وبالتالي يمكن القول بأن الدلالة الإحصائية هي دالة حجم العينة (نصار، ٢٠٠٦). كما أن الدلالة الإحصائية للنتائج البحثية تعتمد أيضا على مستوى الدلالة المتوقع والذي يحدده الباحث من خلال الفرضيات الإحصائية لبحثه، وهنا يجب الانتباه حيث يمكن أن يغير بعض الباحثين قيمة مستوى الدلالة المتوقع تبعا للاحتمال الفعلي للصدفة في نتائجهم Sig or P وبحيث يتم الحصول في النهاية على نتائج دالة إحصائيا. ومن هنا فإنه يمكن القول بأن الاعتماد على الدلالة الإحصائية لوحدها في عملية تفسير النتائج البحثية غير كافي أو حتى أنه قد يكون في بعض الأحيان مضللا.

الدلالة العملية (Practical Significance)

يشير مفهوم الدلالة العملية للنتائج إلى استكشاف قوة العلاقة بين المتغيرات قيد الدراسة من خلال استخدام أساليب إحصائية أقل تأثرا بحجم العينة من تلك الأساليب الإحصائية التي تعتمد على فحص الدلالة الإحصائية للفروق بين المتوسطات أو حتى عند فحص العلاقة الإرتباطية بين المتغيرات. وبصورة أساسية فإنه يمكن الحصول على نتائج دالة إحصائيا –على سبيل المثال- مع عدم وجود فروق كبيرة بين المتوسطات فقط لأن الباحث قد استخدم عينات كبيرة الحجم نسبيا، حيث أنه من الناحية الإحصائية فإن قيمة الخطأ المعياري (Standard Error) تقل كلما ازداد حجم العينة وبالتالي تزداد قوة الاختبار الإحصائي المستخدم لفحص الفرضية أو الفرضيات الصفرية. وبمعنى آخر فإن فرقا مقداره ٥ أو ١٠ أو ٢٠ أو ٣٠ بين المجموعتين التجريبية والضابطة قد يكون كافيا للحصول على نتائج دالة إحصائيا بالرغم أن الدلالة العملية لتلك الفروق ليس بالضرورة أن تكون بنفس المستوى (نصار، ٢٠٠٦). ولذلك فإن هناك الكثير من المؤسسات المسؤولة عن نشر الأبحاث النفسية والتربوية التي لا تقبل نتائج الدراسات التي تخلو من الإشارة إلى الدلالة العملية.. ويعتبر حجم الأثر (Effect Size) من أشهر الأساليب التي تستخدم لفحص الدلالة العملية، وهو أسلوب إحصائي كمي يقيس درجة

أهمية المعالجة من خلال فحص الفرق بين المتوسطات بغض النظر عن الدلالة الإحصائية لتلك الفروق، وبمعنى آخر فقد تكون النتائج دالة إحصائيا، ولكنها قد لا تكون دالة عمليا (Thompson, 2006). هناك عدة أساليب لاستخدام حجم الأثر، فعلى سبيل المثال هناك ما يسمى: حجم الأثر المعدل (Corrected Effect Size) وهناك أيضا أساليب تعتمد على فحص الفروق بين المتوسطات (مثل معادلة كوهين "د") وهناك أساليب تستخدم نسبة التباين الذي يفسره المتغير المستقل من تباين المتغير التابع (مثل مربع إيتا). إن الهدف الأساس من جميع الأساليب السابقة هو الإجابة على السؤال التالي: هل أحدثت المعالجة فرقا عمليا بين المجموعات التجريبية والضابطة، وما مقدار الفرق الذي حققته؟ ويعتبر أسلوب حجم الأثر مفيدا لأنه يهتم بتقرير أهمية النتائج على المستوى العملي وهو بذلك يتعدى مفهوم الدلالة الإحصائية التي تهتم فقط فيما إذا كانت النتائج تعود إلى عامل الصدفة ام لا، ويعتبر مفهوم حجم الأثر مفهوما مهما في الدراسات النفسية والتربوية ، وفي هذا السياق إعتبرت الجمعية النفسية الأمريكية حجم الأثر كأحد العناصر الأساسية التي يجب أن تؤخذ بعين الاعتبار من قبل المجلات النفسية لقبول نشر تلك الدراسات أو رفضها (Wilkison & APA Task Force on statistical Inference, 1999).

الدلالة الإكلينيكية (Clinical Significance): إن معظم الأساليب المستخدمة في تقدير حجم الأثر في الدراسات النفسية والتربوية تعتمد على الفرق بين المجموعات التجريبية والضابطة بدون أن تهتم بذلك التغير الذي يتم على مستوى الفرد الواحد. وهناك اتجاه حديث في البحث النفسي والتربوي يشير إلى ضرورة قياس ورصد مستوى التغيير الحاصل نتيجة للمعالجة على مستوى كل فرد من أفراد المجموعة التجريبية بدلا من التعامل من المتوسطات الحسابية والتي تتعامل معهم كمجموعة. إن الهدف من الدراسات التجريبية في الدراسات النفسية والتربوية هو فحص فيما اذا كانت المعالجة تساعد الناس على التغير نحو الأفضل، سواء كان ذلك في مجال الاضطرابات العقلية أو صعوبات التعلم أو حتى في مجال التشخيص والعلاج، وفي هذا السياق يمكن القول أن فحص تأثير المعالجة او فعاليتها لا يعتبر مهما الا إذا أحدثت التغيير المتوقع والمطلوب سواء في مجال التشخيص أو حتى في مجال العلاج، فعلى سبيل المثال إن الهدف من استخدام برنامج ارشادي او علاجي للتعامل مع مشكلة الاكتئاب لدى مجموعة من الأفراد، هو نقلهم من حالة الاكتئاب إلى حالة من عدم الاكتئاب، وكذلك فإنه يمكن القول بأن فاعلية برنامج لتحسين مهارات القراءة لدى الاطفال تقاس من خلال قدرة ذلك البرنامج على نقل هؤلاء الأطفال من مستوى القراءة الضعيفة إلى مستوى يتوافق مع أقرانهم من نفس المستوى، وبحيث يكمن إعادة دمجهم وإعادتهم الى البرامج العادية. من هنا فإن الأساليب المستخدمة في فحص الدلالة الإكلينيكية لنتائج الدراسات المستخدمة للتصاميم التجريبية تحاول فحص درجة فعالية المعالجة المستخدمة في إحداث التغيير المطلوب للفئة المستهدفة من الأفراد (Campbell, 2005) وهما في المثاليين السابقين: العملاء الذين يعانون من الاكتئاب، والأطفال ذوي المستوى الضعيف في القراءة، ويعتبر "جاكبسون وفولت وريفينس ستورف (Jacobson, Follette, Revenstorf, 1984)" اول من استخدم أسلوبا محددا لفحص الدلالة الإكلينيكية للنتائج. ويعود السبب في اهتمام هؤلاء الباحثين بمفهوم الدلالة الإكلينيكية إلى شعورهم بوجود نقص ما في ميدان الدراسات التي تهتم بمجال العلاج النفسي، حيث أن مجال الاهتمام هنا هو الفرد وليس المجموعة كما هو الحال في الدلالة الاحصائية والدلالة العملية، لقد أشار هؤلاء الباحثين (Jacobson, Follette, Revenstorf, 1984) إلى أن الفرق بين المتوسطات لا يعطي معلومات حقيقية حول عدد المسترشدين الذين انتقلوا من مستوى الاضطراب الى المستوى العادي، ولقد اصبح الأسلوب الذي استخدمه هؤلاء الباحثين أساسا في مجال موضوع الدلالة الإكلينيكية للدراسات النفسية والتربوية، ومنذ

ذلك الوقت ظهرت العديد من الأساليب الأخرى والتي اقترحها العديد من الباحثين لفحص الدلالة الإكلينيكية وبحيث أصبح مفهوم الدلالة الإكلينيكية أحد المؤشرات الرئيسية في تحديد أهمية النتائج ليس على مستوى المجموعات فقط بل أيضا على مستوى الافراد، ولقد أحدث هذا المفهوم نقلة نوعية في مجال البحث النفسي والتربوي تجاوزت في الحقيقة مفهومي أو أسلوبي الدلالة الإحصائية والدلالة العملية (حجم الأثر) اللذان لم يعودا كافيين للإشاره الى اهمية النتائج في الدراسات التجريبية. ولكن هل الدلالة الإكلينيكية والدلالة العملية مفهومان مترادفان أم أن هناك فرق بينهما؟ بما أن مفهوم الدلالة الإحصائية مفهوم جديد نسبيا في مجال الدراسات النفسية والتربوية فإنه يتداخل في كثير من الأحيان مع مفهوم الدلالة العملية للنتائج، حيث يتم في كثير من الأحيان الخلط بين كلا النوعين (Peterson, 2008)، وفي الحقيقة فإن الدلالة الاكلينيكية تختلف تماما عن الدلالة العملية، والمثالين التاليين يوضحان الفارق بين هذين النوعين من الدلالات في سياق الدراسات النفسية والتربوية.

مثال نفسي: لنفرض أنك أخصائي نفسي وتعمل في مدرسة وطلب منك ان تتعامل مع طفل يعاني من مشكلة الاكتئاب، وعند البحث في الأدب وجدت دراستين تقدم كل منهما أسلوبا علاجيا مقترحا لحل مشكلة الاكتئاب لدى الأطفال، كلا الدراستين اختبرت الفرضية الصفرية التالية: لا توجد فروق دالة إحصائيا عند مستوى الدلالة (٠.٠٥)، بين متوسطي درجة الاكتئاب على القياس البعدي للمجموعتين التجريبية والضابطة، لنفرض ان نتائج كل من الدراستين اشارت إلى امكانية رفض تلك الفرضية الصفرية حيث كانت قيمة مستوى الدلالة الفعلي أقل أو تساوي (٠.٠٥)، ولنفرض انها كانت تساوي (٠.٠٣)، في هذه الحالة يمكن الاستنتاج ان كلا من الأسلوبين المستخدمين في علاج مشكلة الاكتئاب لدى الأطفال كانا فعالين، ولنفرض أيضا أن الباحثين في الدراستين اهتموا أيضا بفحص الدلالة العملية للنتائج، ولتحقيق هذا الغرض تم حساب حجم الأثر باستخدام معادلة كوهن "د" (Cohen'S d) ولنفرض أيضا ان كلا الدراستين اشارت أن مقدار حجم الأثر باستخدام المعادلة المشار إليها يساوي (0.9)، وبالرجوع للمعايير التي أشار إليها كوهن، فإنه يمكن وصف تلك القيمة بأنها كبيرة جدا، مما يشير إلى أن الفرق بين متوسطات الدرجات البعدية للمجموعتين: التجريبية، والضابطة في كلتا الدراستين كانت ظاهرة وواضحة وتشير الى وجود دلالة عملية للنتائج في كلا الدراستين. وبالرغم من أن معظم الدراسات النفسية والتربوية تشير في أحسن الأحوال إلى الدلالات الاحصائية والعملية للنتائج، ولكن وللاسف، فإن الدراسات التي تشير الى الدلالة الإكلينيكية تعتبر نادرة جدا، ولربما تكون معدومة على مستوى الوطن العربي. وبالرجوع إلى المثال السابق فإنه يلاحظ ان النتائج كانت دالة إحصائيا في الحالتين، وأنها كانت أيضا دالة عمليا وبنفس المستوى، فهل هذا يعني أنه يمكن للباحث استخدام أي من الأسلوبين على اعتبار انهما يحققان نفس النتائج؟ قبل الإجابة على هذا السؤال، لنفرض ان الباحثين في تلك الدراسات لم يشيروا فقط الى الدلالة الاحصائية والدلالة العملية بل أنهم ايضا أشاروا إلى الدلالة الإكلينيكية في نتائجهم، في هذه الحالة فإنه يمكن ان نجد فروقا بين نتائج كلتا الدراستين، فالمفحوصين في الدراسة الأولى وعلى الرغم من وجود فروق دالة احصائيا وعمليا بين المجموعتين التجريبية والضابطة على القياس البعدي في تلك الدراسة إلا أنهم ما زالوا يعانون من مشكلة الاكتئاب تبعا لدرجاتهم على مقياس معروف في قياس تلك السمة، مثل قائمة (بيك) وهم بالتالي ما زالوا بحاجة إلى العلاج حتى يصلوا الى حالة من عدم الاكتئاب، وفي المقابل إذا أشارت نتائج الدلالة الإكلينيكية في الدراسة الثانية أن ما نسبته (٨٠%) من أفراد تلك الدراسة قد انخفضت درجاتهم البعدية على مقياس الاكتئاب بصورة كافية بحيث انهم انتقلوا الى حالة من عدم الاكتئاب، وبالتالي لم يعودوا بحاجة الى اية معالجة. وبناء على نتائج

الدلالة الإكلينيكية المشار إليها في كلتا الدراستين، فإن الأخصائي النفسي قد يختار الطريقة الثانية لأنه أكثر فاعلية من الناحية الإكلينيكية من الأسلوب المستخدم في الدراسة الأولى.

ثانيا: مثال تربوي، لنفرض أن مدير مدرسة أساسية قد لاحظ أن بعض طلبة الصف الرابع الأساسي يعانون من ضعف مهارة القراءة، لذلك قام بتطبيق اختبار على جميع طلبة الصف الرابع لقياس مهاراتهم في القراءة حيث تم تحديد درجة القطع ٥٠ كلمة في الدقيقة كمعيار للحكم فيما اذا كان الطفل متقنا أو غير متقنا لمهارات القراءة، وفي ضوء ذلك تم تحديد عشرين طفلا كانت درجاتهم في القراءة أقل من ٥٠، وهم بالتالي بحاجة لتدخل المدرسة لتحسين مهاراتهم في القراءة. بعد ذلك، كلف ذلك المدير أحد معلمي اللغة العربية باختيار عشرة أطفال من بين مجموعة أطفال ممن يعانون من ضعف القراءة بحيث يتم توزيعهم إلى مجموعتين: مجموعه تجريبية ومجموعة ضابطة، وبحيث تتعرض المجموعة الأولى لبرنامج تدريبي مكثف لمدة أربع أسابيع يهدف إلى تحسين مهارات الأطفال في القراءة بينما لم تتعرض المجموعة الضابطة لذلك البرنامج، وقد تم إجراء قياس قبلي وقياس بعدي لعدد الكلمات التي يستطيع أطفال كلتا المجموعتين قراءتها قبل وبعد تنفيذ البرنامج، ويشير جدول ١ إلى عدد الكلمات التي استطاع أفراد كلتا المجموعتين قراءتها بطريقة صحيحة خلال دقيقة.

وباستخدام اختبار "ت" للعينات المستقلة لفحص دلالة الفرق بين متوسطي المجموعتين على القياس البعدي عند مستوى الدلالة (٠.٠٥) دلت نتائج هذا التحليل إلى وجود فروق دالة احصائيا بين متوسط عدد الكلمات التي استطاع أطفال المجموعة التجريبية قراءتها بعد البرنامج مقارنة بعدد الكلمات التي تمكن أفراد المجموعة الضابطة من قراءتها خلال دقيقة، كذلك يظهر من خلال المتوسطات الحسابية ان عدد الكلمات التي استطاع طلبة المجموعة التجريبية قراءتها بعد البرنامج كانت أكثر من عدد الكلمات التي استطاع أفراد المجموعة الضابطة قراءتها بعد البرنامج خلال دقيقة. ويوضح الجدول ٢ المتوسطات الحسابية والانحرافات المعيارية لعدد الكلمات التي استطاع أفراد المجموعتين التجريبية والضابطة قراءتها خلال دقيقة بعد البرنامج، كذلك يتضمن ذلك الجدول نتائج اختبار "ت" لفحص دلالة الفروق بين المتوسطين.

جدول ١

عدد الكلمات التي استطاع أفراد الدراسة قراءتها بطريقة صحيحة خلال دقيقه قبل وبعد البرنامج

الاسم	المجموعة	عدد الكلمات: قبل البرنامج	عدد الكلمات: بعد البرنامج
محمد	التجريبية	٤٠	٦٠
عمر	التجريبية	٢٥	٤٧
حمزة	التجريبية	٤٥	٦٥
سوسن	التجريبية	٢٨	٤٦
صفاء	التجريبية	٢٧	٣٢
علي	الضابطة	٢٠	٢٤
قصي	الضابطة	٤٠	٤٦
زين	الضابطة	١٨	٢٣
سلمى	الضابطة	٢٤	٢٠
فاطمة	الضابطة	٣٠	٣٠

جدول ٢

عدد الكلمات التي استطاع أفراد المجموعتين التجريبية والضابطة قراءتها خلال دقيقة بعد البرنامج ونتائج اختبار "ت" لفحص دلالة الفروق بين المتوسطين

المجموعة	العدد	المتوسط الحسابي	الانحراف المعياري	قيمة "ت "	درجات الحرية	مستوى الدلالة
التجريبية	٥	٥٠	١٢.٩٨	٢.٨٨	٨	٠.٠٢١
الضابطة	٥	٢٨.٦	١٠.٣٨			

إن نتائج اختبار "ت" في جدول ٢ تشير إلى وجود فروق دالة احصائيا عن مستوى الدلالة (٠.٠٥) بين المجموعتين التجريبية والضابطة، لكن هذه النتائج لا تعتبر كافية لتحديد درجة استفادة افراد المجموعة التجريبية كأفراد، وليس كمجموعة، من البرنامج وبالتالي فإنها قد لا تكون كافية لاتخاذ القرارات اللازمة لتحسين مهارة القراءة لدى هؤلاء الأفراد. وباستخدام أساليب فحص الدلالة العملية مثل

اختبار كوهين "د" فإننا نحصل على حجم أثر بلغ مقداره (١.٦٤) وتشير هذه القيمة حسب ما اشار ستيفنز (Stevens, 1996) إلى حجم أثر كبير. كما وتشير قيمة مؤشر حجم الأثر هذه إلى أن البرنامج التدريبي قد أحدث فرقا دالا من الناحية العملية وليس فقط من الناحية الإحصائية بين المجموعتين التجريبية والضابطة من حيث عدد الكلمات التي استطاع الأطفال في كلتا المجموعتين قراءتها خلال دقيقة بعد البرنامج. وتكمن أهمية مؤشر حجم الاثر إلى تحديد مقدار الفرق، حيث أنه فرق كبير وواضح. ولفهم فاعلية البرنامج بطريقه أكثر عمقا وأكثر تفصيلا، خاصة من حيث تأثيرها على كل فرد من أفراد المجموعة التجريبية، فإنه من الضروري تقدير الدلالة الإكلينيكية آخذين بعين الاعتبار أن نقطة القطع التي تميز ما بين الأطفال القادرين على القراءة وغير القادرين على القراءة هي (٥٠) خمسين كلمة في الدقيقة، وباستخدام هذا المعيار وبالرجوع للبيانات الواردة في جدول ١ يتضح أن اثنين (محمد، وحمزه) وهما من افراد المجموعة التجريبية لم يتحسنا فقط في القراءة، بل انهما انتقلا من فئة الأطفال غير المتقنين للقراءة إلى فئة الأطفال المتقنين للقراءة، وبينما نجد أن باقي أفراد المجموعة التجريبية (عمر، وسوسن، وصفاء) قد تحسن مستواهم في القراءة، ولكنهم بقوا تحت نقطة القطع المشار إليها، وبمعنى آخر أنهم لم ينتقلوا الى فئة الطلاب المتقنين للقراءة. ومن الملاحظ أيضا من خلال البيانات في جدول ١ ان أفراد المجموعة الضابطة قد حققوا ايضا تحسنا في مستوى القراءة، فكيف يمكن تفسير حدوث ذلك في غياب تعرضهم للبرنامج التدريبي؟ وقد يعود السبب في ارتفاع متوسط درجات أفراد المجموعة الضابطة الى عامل الانحدار نحو المتوسط، فكما هو معروف إحصائيا فإنه عند التعامل مع عينات متطرفة على سمة ما، فإنه يلاحظ ان متوسط القياس البعدي يتجه نحو متوسط الدرجات في تلك السمة، وهذا ما يعرف احصائيا **بالانحدار نحو الوسط.**

طرق تقدير الدلالة الإكلينيكية في الدراسات النفسية والتربوية: هناك عدة طرق لتقدير الدلالة الإكلينيكية وفي الدراسة الحالية تم استعراض أهم خمس طرق مستخدمة في الدراسات النفسية والتربوية لتقدير تلك الدلالة. وهذه الطرق هي طريقة جاكبسون- تروكس (Jacobson-Truax (JT)، وطريقة جوليكسن- لورد- نوفك (Gulliksen-Lord-Novick (GLN)، وطريقة إدواردز- نانولي (Edwards-Nunnally (EN)، وطريقة *هيجمان- أريندال* (Hageman-Arrindell (HA)، وأخيرا الطريقة الخطية الهرمية (Hierarchical Linear Method (HLM) (Bauer, Lambert, & Nielsen, 2004).

ومن الجدير ذكره أنه تم في الدراسة الحالية التركيز فقط على طريقة جاكبسون- تروكس (Jacobson-Truax (JT) لسببين, أولا: لأن هذه الطريقة هي الأكثر استخداما في الدراسات النفسية والتربوية المستخدمة للتصاميم التجريبية وخاصة تلك الدراسات التي يكون الهدف الأساس منها هو إحداث تغيير عند الأفراد نتيجة لتعرضهم لمعالجة محددة، وثانيا: لأن هذه الطريقة- على الرغم من بساطتها- فهي حسب ما أشارت الدراسات تعطي تقديرا متوسطا لأثر المعالجة بمعنى أنها لا تقلل ولا تضخم في نفس الوقت من أثر تلك المعالجة (Bauer et al., 2004).

من جهة ثانية فإن الدراسة الحالية تسعى لتقديم عرضا مختصرا للطرق الأخرى التي تم ذكرها سابقا لتقدير الدلالة الإكلينيكية لنتائج الأبحاث التجريبية حيث أن الهدف من ذلك توفير عدد من الاساليب التي من الممكن استخدامها من قبل الباحثين العرب وخاصة أن هذا المفهوم وعلى الرغم من أنه قد ظهر منذ عام ١٩٨٤ إلا أإنه لم يشر إليه بصورة كافية في الدراسات النفسية والتربوية والتي تم إجراءها في البيئة العربية.

طريقة جاكبسون- تروكس Method-Jacobson-Truax (JT): يعتبر الأسلوب الذي

طوره عام ١٩٨٤ جاكبسون و فوليت وريفنستورف (Jacbson, Follett, & Revenstrof, 1984) هو الطريقة الأساسية لتقدير الدلالة الإكلينيكية في الدراسات النفسية والتربوية. ولقد تم مراجعة هذا الأسلوب من قبل كل من جاكبسون وترواكس (Jacobson and Truax, 1991) عام ١٩٩١ وقد عرف هذا الأسلوب بـــطريقة جي تي (JT Method). ويعتمد هذا الأسلوب على خطوتين أساسيتين حيث يتم في البداية تحديد درجة تسمى بنقطة القطع (Cut point) وهي درجة في المقياس من المفترض أنها تكون قادرة على أن تميز بين الأفراد الذين يحتاجون لمعالجة (Dysfunctional group) والأفراد العاديين أو الذين لا يحتاجون لنفس المعالجة (Functional group)، بينما يتم في الخطوة الثانية تقدير ما يسمى بمؤشر ثبات التغير (Reliability Chang Index) ويرمز له بالرمز (RCI) وفي ضوء هذين المعيارين يتم تصنيف أفراد عينة الدراسات إلى أربعة فئات: فئة الأفراد المعالجين (الذين تمت معالجتهم) (Recovered group)، وفئة الأفراد الذين تحسنوا (Improved group)، وفئة الأفراد الذين لم يحدث لهم أي تغير (Unchanged group)، وفئة الأفراد الذين أصابهم نوع من التدهور أو التراجع على السمة المقاسة (Deteriorated group).

بالنسبة لطرق تقدير نقطة القطع في هذا الأسلوب فإن هناك ثلاث أساليب لتحقيق ذلك حسب ما يشير بيترسون (Peterson, 2008) ويشار في الأدب إلى هذه الطرق بما يسمى بالطريقة أ (A) والطريقة ب (B) والطريقة ج (C). وتبعا للطريقة أ فإنه يمكن تحديد درجة القطع من خلال استخراج تلك الدرجة التي تنحرف بمقدار درجتين معياريتين عن متوسط الدرجات على القياس القبلي لمجتمع الأفراد الذين ما زالو بحاجة لمعالجة (Bauer, Lambert, & Nielsen, 2004).

أ= متوسط الدرجات القبلية (لمجتمع الأفراد الذين هم بحاجة لمعالجة) + ٢ × الانحراف المعياري لدرجاتهم القبلية،

وبمعنى آخر، فإنه يمكن استخراج نقطة القطع أ أو A من خلال حساب المتوسط الحسابي والانحراف المعياري للدرجات القبلية لمجتمع الأفراد الذين هم بحاجة لمعالجة مقدرا من خلال العينة ثم يتم جمع المتوسط الحسابي مع ضعف قيمة الانحراف المعياري لدرجات نفس المجتمع، وفي هذا السياق يشير بيترسون (Peterson, 2008) إلى أن نقطة القطع أ هي درجة في القياس البعدي يجب أن تكون خارج مدى درجات المجتمع الذي تمثله المجموعة التي ما زالت بحاجة للمعالجة. ويشير نفس الباحث إلى أن الدرجة ب أو B هي درجة في القياس البعدي يجب ان تقع ضمن مدى درجات المجتمع الذي تمثله المجموعة المعالجة (التي لا تحتاج إلى معالجة). ويمكن حساب نقطة القطع ب من خلال المعادلة التالية:

ب = متوسط درجات مجتمع العاديين (لا يحتاجون لمعالجة) + ٢ × الانحراف المعياري لدرجات نفس المجتمع

ومن الوضح أن الطريقة ب تفترض توفر معلومات حول توزيع درجات الأفراد العاديين على السمة قيد الدراسة وفي حال عدم توفر مثل هذه المعلومات فإنه لا يمكن استخدام هذه الطريقة لتحديد نقطة القطع، كما أنه من السهل نسبيا على معظم المسترشدين تجاوز هذه النقطة وذلك بسبب أنه في كثير من الأحيان يحصل تداخل (Overlap) بين توزيع درجات مجتمع الأفراد الذين لا يحتاجون للمعالجة (العاديين) وتوزيع درجات مجتمع الأفراد الذين يحتاجون للمعالجة Bauer, Lambert, & Nielsen, (2004).

أما بالنسبة لنقطة القطع ج أو C فهي تلك الدرجة البعدية التي يجب أن تكون أقرب لمتوسط درجات مجتمع الأفراد العاديين من متوسط درجات مجتمع الأفراد الذين هم بحاجة لمعالجة (Peterson, 2008)، أو هي تلك الدرجة التي من المتوقع أن تقع بين متوسط مجتمع العاديين ومجتمع الأفراد الذين بحاجة للمعالجة.

ويمكن حساب هذه النقطة باستخدام المعادلة التالية (Bauer, Lambert, & Nielsen, 2004):

ج = [(ع بحاجة للمعالجة × م العاديين) + (ع العاديين × م بحاجة للمعالجة)] / (ع بحاجة للمعالجة + ع العاديين)

حيث: ع بحاجة للمعالجة: الانحراف المعياري لدرجات مجتمع الأفراد الذين بحاجة للمعالجة، م العاديين: متوسط درجات مجتمع الافراد العاديين (الذين لا يحتاجون للمعالجة)، ع العاديين: الانحراف المعياري لدرجات مجتمع الافراد العاديين (الذين لا يحتاجون للمعالجة)، م بحاجة للمعالجة: متوسط درجات مجتمع الأفراد الذين بحاجة للمعالجة.

وتعتبر الطريقة السابقة في تقدير نقطة القطع هي الأسلوب الأفضل في حال توفر معلومات عن توزيع درجات مجتمع الأفراد العاديين (الذين لا يحتاجون للمعالجة) وتوزيع درجات مجتمع الأفراد الذين يحتاجون للمعالجة وخاصة إذا كان هناك تداخلا بين التوزيعين (Bauer, Lambert, & Nielsen, 2004). كما أن هذه الطريقة في تقدير نقطة القطع تعتبر أكثر دقة من الطريقتين السابقتين خاصة أنها تعتمد على الاحتمالية النسبية لدرجة ما يتم الحصول عليها من خلال مجتمع محدد في مقابل مجتمع آخر وبمعنى آخر فإن هذه الدرجة والتي يتم الحصول عليها من خلال الخصائص الإحصائية لتوزيع درجات مجتمع الأفراد العاديين وتوزيع درجات مجتمع الأفراد الذين هم بحاجة للمعالجة، مما يجعلها نسبيا الأكثر دقة مقارنة بالطريقتين المشار إليهما سابقا (Jacobson, Roberts, Berns, & McGlinchey, 1999). وفي المقابل يعتبر استخدام الطريقتين أ و ب أو A و B أكثر منطقية في حال عدم توفر البيانات الكافية لحساب نقطة القطع حسب الطريقة ج أو C، ومن الجدير ملاحظته أن الطريقة أ أو A تعتبر هي الأفضل في حال عدم توفر بيانات كافية حول طبيعة توزيع درجات مجتمع الأفراد العاديين ومجتمع الأفراد الذين بحاجة للمعالجة. وبعد أن يتم تحديد نقطة القطع وفقا لأي من الأساليب السابقة، فإن الخطوة التالية هي تحديد مقدار التغير الذي حدث وذلك من خلال استخدام ما يسمى بمؤشر ثبات التغير(RCI)- (Reliability Change Index)، ويتم حساب قيمة هذا المؤشر لكل فرد على حدة وذلك من خلال حساب الفرق بين درجاته على القياسين القبلي والبعدي ومن ثم قسمة هذا الفرق على الخطأ المعياري لدرجات الفرق (محسوبا من خلال جميع أفراد الدراسة)، ووفقا لهذا الأسلوب فإن قيمة مؤشر ثبات التغير قد تكون موجبة (إذا كانت الدرجة على القياس البعدي أكبر من الدرجة على القياس القبلي) أو صفر (إذا كانت الدرجة على القياس القبلي مساوية للدرجة على القياس البعدي) أو سالبة (إذا كانت الدرجة على القياس البعدي أصغر من الدرجة على القياس القبلي).

ومن خلال استخدام درجة القطع ومؤشر ثبات التغير(RCI) فإنه يتم تصنيف أفراد عينة الدراسة إلى أربع فئات وهي: فئة الأفراد المعافيين (Recovered): وهم أولئك الأفراد الذين وصلوا لحالة الشفاء ولم يعودوا بحاجة لاي معالجة، ويعرفوا وفقا لطريقة طريقة جاكبسون-تراوكس (JT) بأنهم أولئك الأفراد الذين تجاوزوا نقطة القطع المحددة وحصلوا على مؤشر ثبات تغير ذو قيمة موجبة، وفئة الأفراد الذين تحسنوا (Improved): وهم أولئك الأفراد الذين تحسنت دجاتهم البعدية مقارنة بدرجاتهم القبلية ولكنهم ما زالوا بحاجة للمعالجة، ويعرفوا وفقا لطريقة طريقة جاكبسون-تراوكس (JT) بأنهم أولئك الأفراد الذين لم يتجاوزوا نقطة القطع المحددة ولكنهم حصلوا على مؤشر ثبات تغير ذو قيمة موجبة، وفئة الأفراد الذين لم يتغيروا (Unchanged): وهم أولئك الأفراد الذين لم يحدث أي تغير لديهم وبالتالي درجاتهم ما زالوا بحاجة للمعالجة، ويعرفوا وفقا لطريقة طريقة جاكبسون-تراوكس (JT) بأنهم أولئك الأفراد الذين لم حصلوا على مؤشر ثبات تغير يساوي صفر نتيجة لتساوي درجاتهم القبلية مع درجاتهم البعدية، وفئة الأفراد الذين تدهوروا أو تراجعوا (Deteriorated): وهم أولئك الأفراد الذين كانت درجاتهم البعدية أقل من درجاتهم القبلية

وما زالوا بحاجة للمعالجة، ويعرفوا وفقا لطريقة طريقة جاكبسون- تروكس (JT) بأنهم أولئك الأفراد الذين لم يتجاوزوا نقطة القطع المحددة وحصلوا على مؤشر ثبات تغير ذو قيمة سالبة.

ولتوضيح كيفية تقدير الدلالة الإكلينيكية وفقا لطريقة طريقة جاكبسون- تروكس (JT) فإنه سيتم الرجوع إلى البيانات المشار إليها في المثال التربوي (جدول ١) والمتعلقة بفحص فعالية برنامج تدريبي في تحسين مهارات القراءة لدى أطفال الصف الرابع، حيث تم حساب مؤشر ثبات التغير (RCI) لكل فرد من أفراد المجموعتين التجريبية والضابطة وباستخدام ذلك المؤشر ونقطة القطع ٥٠ كلمة في الدقيقة تم تصنيف كل فرد من أفراد الدراسة الافتراضية السابقة إلى إحدى الفئات الأربع المذكورة سابقا ويشير جدول ٣ إلى تلك النتائج.

جدول ٣

تصنيف أفراد العينة في الدراسة الافتراضية والمتعلقة بتحسين مهارات القراءة لدى أطفال الصف الرابع تبعا لطريقة جاكبسون- تروكس (JT)

الاسم	المجموعة	عدد الكلمات في القياس البعدي	مؤشر ثبات التغير (RCI)	التصنيف أو وصف التغير
محمد	التجريبية	٦٠	٦.٨٩	تعالج
عمر	التجريبية	٤٧	٧.٥٨	تحسن
حمزة	التجريبية	٦٥	٦.٨٩	تعالج
سوسن	التجريبية	٤٦	٦.٢٠	تحسنت
صفاء	التجريبية	٣٢	١.٧٢	تحسنت
علي	الضابطة	٢٤	١.٣٨	تحسن
قصي	الضابطة	٤٦	٢.٠٧	تحسن
زين	الضابطة	٢٣	١.٧٢	تحسن
سلمى	الضابطة	٢٠	-١.٣٨	تراجع
فاطمة	الضابطة	٣٠	٠.٠٠	ثبت أو لم يتغير

يتبين من خلال جدول ٣ أن طفلين من المجموعة التجريبية قد إنتقلوا إلى مستوى الشفاء أو التعافي (Recovery Stage)، حيث يلاحظ أن الطفلين محمد وحمزة قد تجاوزا نقطة القطع المحددة وهي ٥٠ كلمة في الدقيقة وكان مؤشر ثبات التغير لديهما موجبا. أما الأطفال عمر وسوسن وصفاء - وهم باقي أفراد المجموعة التجريبية- فقد صنفوا إلى فئة المتحسنين وكان أساس هذا التصنيف أن كل منهم قد حصل على مؤشر ثبات تغير موجب ولكنه فشل في تجاوز نقطة اقطع من حيث عدد الكلمات التي استطاع قراءتها بطريقة صحيحة خلال دقيقة في القياس البعدي.

أما بالنسبة لتصنيف أفراد المجموعة الضابطة، فكما يتضح في جدول ٣ فإن ثلاثة من أفرادها وهم علي وقصي وزين قد صنفوا أيضا إلى فئة المتحسنين ومرة أخرى فإنه يمكن تبرير ذلك التحسن من خلال مبدأ إحصائي معروف وهو انحدار القيم نحو المتوسط، ولتوضيح ها المبدأ لنفرض أن ١٠٠ طالب قد تقدموا لإختبار مكون من ١٠٠ فقرة من نوع الإختيار من متعدد في مادة الإحصاء حيث كان متوسط درجاتهم في هذا الإختبار يساوي ٧٠ درجة، ولنفرض أن مدى الدرجات على ذلك قد تراوح من ٣٠ إلى ٩٠ وبحيث أن خمسة طلاب أي ما نسبته ٥% من الطلبة قد حصلوا على درجات متدنية جدا ٣٠-٤٠، وفي المقابل فإن نفس النسبة من الطلبة (٥ طلاب)، حصلوا على درجات مرتفعه ٨٠-٩٠، فإنه من المتوقع إحصائيا في هذه حالة تطبيق اختبار مكافئ للاختبار الأول على نفس الأفراد بعد فترة زمنية أن تزداد درجات طلبة الفئة الأولى وأن تقل درجات الفئة الثانية بحبث تقترب درجات طلبة هاتين الفئتين من متوسط الدرجات (٧٠)، وتعرف هذه الظاهرة إحصائيا "بإنحدار القيم نحو المتوسط".

وبالرجوع إلى نتائج جدول ٣ يلاحظ أن الطفلة فاطمة- وهي من افراد المجموعة الضابطة- قد صنفت إلى فئة الثابتين (لم يحصل لديهم تغير في السمة المقاسة) وكان أساس هذا التصنيف أن درجة هذه الطالبة على القياس القبلي قد تساوت مع درجتها على القياس البعدي (انظر بيانات جدول ١) وأن قيمة مؤشر ثبات التغير في أدائها كان يساوي صفر. وأما بالنسبة للطفلة سلمى –وهي أيضا من المجموعة الضابطة- قد صنفت إلى فئة المتراجعين بما أن

درجتها على القياس البعدي كانت اقل من درجتها على القياس القبلي (انظر بيانات جدول ١) كما أن مؤشر ثبات التغير لدى هذه الطالبة كان سالبا.

طريقة جوليكسن- لورد- نوفك (Gulliksen-Lord-Novick (GLN): تم اقتراح هذه الطريقة لتقدير الدلالة الاكلينيكية للنتائج كمحاولة لتجاوز أو لتصحيح الخطأ الذي وقعت به طريقة جاكبسون- ترواكس - (Jacobson-Truax (JT) وذلك من وجهة نظر بعض الاحصائين مثل العالم هاسو (Hsu,1999)، حيث أشار هاسو أن استخدام هذه الطريقة للفرق بين الدرجات القبلية والدرجات البعدية في تحديد قيمة معامل ثبات التغير (RCI) لم يراعي ويضع بعين الاعتبار احتمال انحدار القيم نحو الوسط، ولحل هذه المشكلة ابتكر هاسو طريقة جديدة لتقدير الدلالة الإكلينيكية اعتمدت بصورة أساسية على بعض الطرق الإحصائية والتي قدمها العالم جوليكسين (Gulliksen, 1950) في عام ١٩٥٠ والعالمان لورد ونوفك (Lord and Novic, 1968) في عام ١٩٦٨ ومن هنا يتم استخدام الاختصار (GLN) عند الإشارة لهذه الطريقة في تقدير الدلالة الإكلينيكية. ووفقا لطريقة (GLN) فإنه يجب في البداية استخدام متوسط حسابي لمجتمع افتراضي وبحيث يتم طرح متوسط الدرجات القبلية والبعدية لكل فرد من أفراد المجموعة التجريبية من ذلك المتوسط وذلك كأسلوب لحل مشكلة انحدار القيم نحو الوسط، وكذلك فإنه وفقا لهذه الطريقة (GLN method) فإنه يجب قسمة الفرق بين القياسين القبلي والبعدي على الانحراف المعياري لدرجات المجتمع الإفتراضي بدلا من قسمة الفارق على الخطأ المعياري كما هو الحال تبعا لطريقة جاكبسون- ترواكس(Peterson, 2008).

وطريقة إدواردز- نانولي Edwards-Nunnally (EN): إنتقد العالم سبير (Speer, 1992) طريقة جاكبسون- ترواكس (JT) لنفس السبب الذي أشار إليه هاسو (Hsu, 1999) وهو أن هذه الطريقة قد تتأثر بإنحدار القيم نحو الوسط، وبناء على ذلك قدم سبير(١٩٩٢) طرقة أخرى لتقدير الدلالة الاكلينيكية تعتمد على الأفكار التي قدمها كل من إدواردز (Edwards, 1978) ونانولي (Nunnally,1965). وتقوم فكرة هذه الطريقة بصورة أساسية على تعديل الدرجات القبلية وبحيث تقترب من متوسط الدرجات القبلية وبمعنى آخر فإن الدرجة القبلية للفرد بعد التعديل تقترب بصورة أكبر من متوسط الدرجات القبلية وبالتالي تصبح الدرجات القبيلية أكثر تجانسا وبالتالي يقل تأثير عامل انحدار القيم نحو الوسط خاصة في حالة القيم المتطرفة، أما الخطوة التالية تبعا لطريقة إدواردز- نانولي (EN) فتتمثل بتقدير التغير في أداء الفرد وذلك بإستخدام طريقة فترة الثقة (Confidence Interval method) اعتمادا على الدرجة القبلية المعدلة بدلا من الدرجة الملاحظة. وهنا تشير بيترسون (Peterson, 2008) إلى أن استخدام أسلوب فترة الثقة يتطلب فرقا أكبر بين القياسين القبلي والبعدي للفرد من ذلك الفرق والذي يتم الحصول عليه من خلال طريقة جاكبسون- ترواكس حتى يعتبر دال إكلينيكيا.

طريقة هيجمان- آريندال (Hageman-Arrindell (HA): إقترح كل من هيجمان وآريندل (Hageman and Arrindell, 1999)، ضرورة القيام بإجراء تعديلين أساسيين على الطريقة التي قدمها جاكبسون و ترواكس (Jacobson and Truax- (JT), 1991)، ويتمثل التعديل الأول بضرورة استخدام أساليب إحصائية مختلفة للتمييز بين التغير على مستوى الفرد (CSindiv) والتغير على مستوى المجموعة (CSgroup) بصورة أكثر وضوحا. أما التعديل الثاني فيتضمن الإتفاق مع هاسو(Hsu, 1999) وسبير(Speer, 1992) فيما يتعلق بوجوب تعديل معادلة (JT) بحيث يتم الأخذ بعين الاعتبار مشكلة إنحدار القيم نحو الوسط. ولحل هاتين المشكلتين اقترح كل من هيجمان وآرينديل(Hageman and Arrindell, 1999) استخدام مؤشرين إحصائيين جديدين تم

تطويرهما من قبل كرونباخ (Cronbach & Gleser, 1959) عام ١٩٥٩. والمؤشران المقترحان هما مؤشر ثبات التغير للفرد (RCindiv)، وفيه من الضروري تصنيف الفرد بنسبة دقة لا تقل عن ٩٥%. وفي نفس السياق اشار كل من هيجمان وآريندل (١٩٩٩)، بأنه يمكن تفسير الدلالة الإكلينيكية للتغير على مستوى الفرد (CSindiv) من خلال تعديل طريقة حساب نقطة القطع والمستخدمة في طريقة جاكبسون و ترواكس (Jacobson and Truax- (JT), 1991) وذلك عن طريق الأخذ بعين الإعتبار الدرجة الحقيقية ومعامل ثبات التغير لكل فرد وبحيث يتم استخدام الدلالة الإكلينيكية للفرد لتصنيفه في واحدة من المجموعات التالية: المتدهورين أو المتراجعيين (Deteriorated) أو مجموعة من الأفراد غير ثابتي التغير (Not reliably changed) أو مجموعة الأفراد المتحسنين ولكنهم لم يتعالجوا (Improved but not recovered)، أو مجموعة من المستفيدين المعالجين أو الذين وصلوا إلى مرحلة العلاج (Recovered). ويعتبر مؤشري التغير على مستوى المجموعة (RCgroup) والدلالة الإكلينيكية للمجموعة (CSgroup) من أهم المفاهيم التي قدمها كل من هيجمان وآريندل (Hageman and Arrindell, 1999) في موضوع الدلالة الإكلينيكية وقد اقترحا أساليب إحصائية لحساب كل مؤشر.

الطريقة الخطية الهرمية (Hierarchical Linear Method (HLM)): تم تقديم هذا الأسلوب لتقدير الدلالة الإكلينيكية من قبل سبير وجرينبوم (Speer and Greenbaum, 1995) في عام ١٩٩ (Bauer, Lambert, & Nielsen, 2004). ويعتمد هذا الأسلوب بصورة أساسية على نماذج منحنى النمو (Growth curve models) وليس على الفرق بين القياس القبلي والقياس البعدي كما هو الحال في الطرق التي سبق ذكرها. ومن الضروري في هذا الأسلوب الحصول على ثلاث قياسات على الأقل للفرد في مراحل مختلفة، وبعد ذلك يتم استخدام معادلات خاصة لتحديد درجة التغير على مستوى الفرد وفي الحقيقية إن الحسابات المستخدمة في هذا الأسلوب صعبة نسبيا لذلك فإن هناك برامج إحصائية خاصة من الضروري استخدامها لتقدير الدلالة الإكلينيكية وفقا لهذا الأسلوب والذي تعتبره بيترسون (Peterson, 2008) من أكثر الأساليب مرونة وفائدة لتقدير الدلالة الإكلينيكية مقارنة بالطرق التقليدية الأخرى. وأخيرا تجدر الإشارة إلى أن هذه الطريقة تحتاج لبحث معمق قد يكون خارج إطار أهداف البحث الحالي.

نموذج مقترح لفحص الدلالة الإكلينيكية للمعالجات التجريبية على مستوى المجموعات

ولتحقيق الهدف الأخير من أهداف الدراسة الحالية فإن النموذج المقترح التالي يهدف إلى تحديد الدلالة الإكلينيكية للمعالجات على المجموعات التجريبة والضابطة ، ويعتمد هذا النموذج على استخدام الدلالة الإحصائية ولكن بعد تصنيف أفراد المجموعتين التجريبية والضابطة وفقا للمعايير المستخدمة في طريقة جاكبسون- ترواكس (Jacobson and Truax- (JT), 1991) لتقدير الدلالة الإكلينيكية، ومن ثم استخدام مربع كاي لجداول التوافق (Chi- Square for crosstabulation) لفحص فيما إذا كان هناك علاقة دالة إحصائيا بين متغيري المجموعة (تجريبية – ضابطة) وتصنيف أفراد العينة تبعا للطريقة المذكورة، وتم استخدام البيانات الافتراضية الواردة في المثال التربوي المشار إليه سابقا في سياق الدراسة الحالية لتبيان أسلوب تطبيق النموذج المقترح. وبالرجوع إلى النتائج الواردة في جدول ٣ يلاحظ أن توزيع أعداد أفراد المجموعتين التجريبية والضابطة تبعا لتصنيفهم حسب طريقة جاكبسون- ترواكس(JT) قد كان حسب كما هو موضح في جدول ٤.

يتضح من خلال جدول ٤ أن إثنان من أفراد المجموعة التجريبية قد وصلوا إلى مرحلة العلاج وأن باقي أفراد المجموعة (ثلاثة أفراد) قد تحسنوا، أما بالنسبة للمجموعة الضابطة فإنه لم يصل إي فرد فيها لمرحلة العلاج وهذه

نتيجة متوقعة نتيجة لغياب المعالجة، أما باقي أفراد المجموعة الضابطة فقد توزعوا على باقي المستويات وذلك على النحو التالي: ثلاثة افراد ضمن فئة المتحسنين و وفرد واحد ضمن فئة الثابتين وفرد واحد ضمن فئة المتراجعين.

جدول ٤

توزيع أعداد أفراد العينة في الدراسة الإفتراضية والمتعلقة بتحسين مهارات القراءة لدى أطفال الصف الرابع تبعا لطريقة جاكبسون- تراواكس (JT)

التصنيف المجموعة	المعالجيين	المتحسنين	الثابتين (لم يتغيروا)	المتدهورين أو المتراجعين
التجريبية	٢	٣	صفر	صفر
الضابطة	صفر	٣	١	١

ولفحص الدلالة الإكلينيكية للمعالجة على مستوى المجموعات فقد تم إستخدام إختبار مربع كاي لجدوال التوافق (Chi- Square for crosstabulation) لفحص فيما إذا كان هناك علاقة دالة إحصائيا بين متغيري المجموعة والتصنيف. وقد اشارت نتائج هذا التحليل إلى أن قيمة مربع كاي قد بلغت ٤ وهي غير دالة إحصائيا ($\alpha > 0.05$) مما يعني عدم وجود علاقة دالة بين متغيري المجموعة والتصنيف. وهذا يدل على أن المعالجة المستخدمة في الدراسة الافتراضية لم تكن ذات تاثير كافي بحيث تصل إلى مستوى الدلالة الإكلينيكية على مستوى المجموعات، على الرغم من ان نفس البيانات كانت كافية للحصول على دلالة إحصائية ودلالة عملية للفروق بين المجموعتين التجريبية والضابطة.

لكن ماذا لو تم افتراض أن توزيع افراد المجموعتين كان مختلفا وبحيث كان أثر المعالجة أكثر وضوحا وذلك تبعا للأعداد الواردة في جدول ٥، فهل ستختلف في هذه الحالة نتائج مربع كاي؟

ولفحص العلاقة بين متغيري المجموعة والتصنيف تبعا لتوزيع افراد المجموعتين التجريبية والضابطة على الفئات الأربع حسب البيانات التي تظهر في جدول ٥ وفقا لطريقة جاكبسون- تراواكس (JT)، فقد تم استخدام اختبار مربع كاي لجدوال التوافق لتحقيق الهدف السابق. وقد دلت نتائج هذا التحليل على أن قيمة مربع كاي قد بلغت ٨ وهي دالة إحصائيا ($\alpha \leq 0.05$) مما يشير إلى أن هناك دلالة إكلينيكية لأثر المعالجة على مستوى المجموعات، وبمعنى آخر أن النتائج اختلفت من حيث اللدلالة الإكلينيكية على مستوى المجموعات مقارنة بالنتائج المعتمدة على البيانات الافتراضية في جدول ٤ وذلك لأن أثر المعالجة كان أكثر وضوحا في المثال الإفتراضي الثاني.

جدول ٥

توزيع افتراضي لتوزيع أعداد أفراد مجموعتين تجريبية وضابطة تبعا لطريقة جاكبسون- تراواكس (JT)

التصنيف المجموعة	المعالجيين	المتحسنين	الثابتين (لم يتغيروا)	المتدهورين أو المتراجعين
التجريبية	٤	١	صفر	صفر
الضابطة	صفر	١	٣	١

يظهر من خلال النتائج السابقة أنه عند استخدام الأسلوب أو النموذج المقترح من خلال الدراسة الحالية لفحص الدلالة الإكلينيكية للنتائج على مستوى المجموعات، فإنه يمكن القول أنه عندما تكون النتائج دالة إكلينيكيا على مستوى المجموعات فهي بالضرورة ستكون دالة من الناحيتين الإحصائية والعملية ولكن العكس غير صحيح.

التوصيات

إن الهدف الأساس من الدراسة الحالية هو تقديم مفهوم الدلالة الإكلينيكية للباحثين في مجال العلوم النفسية والتربوية وحتى في كافة ميادين العلوم الإنسانية، كذلك هدفت الدراسة إلى التمييز بين الدلالة الإحصائية والدلالة العملية من جهة والدلالة الإكلينيكية من جهة أخرى. كما حاولت الدراسة تقديم أساليب إحصائية محددة لفحص الدلالة الإكلينيكية لنتائج الأبحاث النفسية والتربوية والتي استخدمت المنهج التجريبي لتحقيق أهدافها أو للإجابة على اسئلتها البحثية. كما ركزت الدراسة الحالية على أن أسلوب الدلالة

الإكلينيكية هو خاص برصد فيما إذا كان التغير الذي يحصل للفرد بعد التجربة كافيا للإدعاء بأنه قد انتقل من حالة المرض أو المعاناة من مشكلة سلوكية أو نفسية أو تربوية إلى الشفاء أو التحسن، وربما في بعض الأحيان قد لا يكون هذا التغير كافيا لإطلاق تلك الإدعاءات بل ربما قد يكون هذا التغير غير ثابت وهو مايطلق عليه من قبل بعض الأساليب التي تم إستعراضها في سياق الدراسة الحالية "بحالة عدم ثبات التغير" (No Reliable change) أو حتى في أسوء الحالات قد تتدهور او تتراجع حالة الفرد بعد المعالجة عما كانت عليه قبل المعالجة. وقد استعرضت الدراسة الحالية بعض الأساليب الإحصائية لفحص الدلالة الإكلينيكية مع أن الباحث قد قرر التركيز على طريقة جاكبسون- تراواكس (Jacobson and Truax- (JT), 1991)، وذلك لأن هذه الطريقة هي الأكثر استخداما في الدراسات النفسية والتربوية المستخدمة للتصاميم التجريبية، ولأن هذه الطريقة- على الرغم من بساطتها- فهي حسب ما أشارت الدراسات- تعطي تقديرا متوسطا لأثر المعالجة، بمعنى أنها لا تقلل ولا تضخم في نفس الوقت من أثر تلك المعالجة (Bauer, Lambert, & Nielsen, 2004)، كما أن بساطة الأساليب الإحصائية التي يمكن استخدامها في هذه الطريقة يجعل من إمكانية التوصية باستخدامها من قبل الباحثين أمرا متاحا سواء كانت معرفتهم أو مهاراتهم الإحصائية متقدمة أو عادية، حيث يمكن وبسهولة توظيف بعض البرامج الإحصائية المعروفة مثل برنامج الحزمة الإحصائية للعلوم الإنسانية (SPSS) لاستخراج القيم الإحصائية التي تتطلبها هذه الطريقة لتحديد الدلالة الإكلينيكية للنتائج.

ومن الجدير ذكره أنه يمكن للباحثين المهتمين بإستخدام الطرق الأخرى والمذكورة أو حتى غير المذكورة في سياق الدراسة الحالية الرجوع إلى بعض المراجع أو الدراسات- والتي تم الإشارة إلى بعضها في سياق الدراسة الحالية- والتي تمكنهم من تحقيق أهداف أبحاثهم. وفي نفس السياق فإنه يمكن القول بأن بعض المفاهيم الإحصائية المهمة قد لا يكون من الممكن توظيفها بصورة عملية وموسعة في الأبحاث النفسية والتربوية من قبل الباحثين في تلك الميادين نتيجة لصعوبة المعادلات الإحصائية الواجب إستخدامها في تلك الاساليب للخروج بدلالات ومعاني بحثية محددة، من هنا فإنه من الأفضل – من وجهة نظر الباحث- توفير الاساليب الإحصائية البسيطة والتي من الضروري بمكان أن تكون مناسبة ومقبولة من الناحية العلمية لإنجاز الاهاف البحثية المقصودة. ومن خلال ما تم ذكره كان تركيز الدراسة الحالية على طريقة جاكبسون- تراواكس (JT) دون غيرها من الطرق مع الإشارة إلا أهمية الطرق الأخرى خاصة أنه تم طرحها لتفادي بعض المشكلات أو العيوب في طريقة جاكبسون- تراواكس (JT) وخاصة مشكلة إنحدار القيم نحو الوسط. كما أن أحد أهم أهداف الدراسة الحالية كان تقديم نموذج مقترح من قبل الباحث لفحص الدلالة الإكلينيكية على مستوى المجموعات وليس فقط على مستوى الأفراد كما هو الحال في جميع الطرق التي تم إستعراضها في سياق الدراسة الحالية، ولتحقيق هذا الهدف تم توظيف طريقة جاكبسون- تراواكس (JT) لتحديد أو لتصنيف أفراد عينة الدراسة في المجموعات التجريبية والضابطة إلى إحدى الفئات الأربع والتي ارتبطت بتلك الطريقة وهذه الفئات هي: المعالجين، والمتحسنين، والثابتين، والمتراجعين، وفي خطوة لاحقة تم إقتراح إستخدام إختبار مربع كاي لجدوال التوافق (Chi- Square for crosstabulation) لفحص فيما إذا كان هناك علاقة دالة إحصائيا بين متغيري المجموعة والتصنيف، ودلت النتائج التي تم الحصول عليها من خلال تحليل بيانات افتراضية قدرة هذا النموذج المقترح على فحص الدلالة الإكلينيكية – من وجهة نظر الباحث- على مستوى المجموعات بطريقة فعالة ومقبولة علميا وعمليا.

وأخيرا فإن إن مشكلة الأساليب الإحصائية والمستخدمة على نطاق واسع في الدراسات

النفسية والتربوية أنها تتعامل مع المتوسطات والانحرافات المعيارية وغيرها من القيم الإحصائية والتي يتم إستخراجها من خلال المجموعات وهنا قد يتم – وبصورة غير مقصودة – إغفال ذلك التغير الذي من المفترض أن يحصل على مستوى الفرد وليس على مستوى المجموعة وبحيث يتم رصد ذلك التغير الذي حصل للفرد بعد تعرضه للمعالجة حيث من المفترض أن ينتقل من حالة المرض او المعاناة من مشكلة سلوكية أو انفعالية إلى حالة الشفاء أو على الأقل التحسن. إن مراجعة سريعة للأبحاث النفسية والتربوية والتي استخدمت التصاميم التجريبية يمكن من ملاحظة أن معظم هذه الدراسات إهتمت بالدلالة الإحصائية والقليل منها إهتم بالدلالة العملية للنتائج والاقل الاقل أشار إلى ذلك المفهوم الهام جدا ألا وهو الدلالة الإكلينيكية. وأخيرا فإنه من المهم التوصية بضرورة فحص الدلالة الإكلينيكية لنتائج الابحاث النفسية والتربوية والمستخدمة للتصاميم التجريبية سواء على مستوى الأفراد أو على مستوى المجموعات وعدم الإكتفاء بالدلالة الإحصائية أو حتى الدلالة العملية، فكما أظهرت الدراسة الحالية فإنه قد تكون النتائج البحثية التجريبية دالة إحصائيا وعمليا ولكنها ليس بالضرورة أن تكون دالة من الناحية الإكلينيكية خاصة على مستوى الأفراد.

المراجع
References

الثبيتي، علي بن حامد (٢٠٠٨). تصاميم البحوث العلمية ودورها في صدق نتائج الدراسات التربوية. **مجلة رسالة الخليج العربي، العدد ١٠٨**، السنة ٢٩، ١-١١٣.

الصياد، عبد العاطي (١٩٨٨). **الدلالة العملية وحجم العينة المصاحبتين للدلالة الاحصائية لإختبار t في البحث النفسي والتربوي (دراسة تقويمية).** جامعة الزقازيق، جمهورية مصر العربية.

عودة احمد والخليلي، خليل (١٩٨٨). **الاحصاء للباحث في التربية والعلوم الانسانية**، عمان: دار الفكر للنشر والتوزيع.

النجار، عبدالله (١٩٩٠). دراسة تقويمية مقارنة للاساليب الاحصائية التي أستخدمت في تحليل البيانات رسائل الماجستير في كل من كلية التربية في جامعة أم القرى بمكة المكرمة وكلية التربية في جامعة الملك سعود بالرياض. **رسالة ماجستير غير منشورة**، كلية التربية، جامعة أم القرى.

نصار، يحيى (٢٠٠٦). استخدام حجم الاثر لفحص الدلالة العلميه للنتائج في الدراسات التربويه والنفسيه المستخدمه للمنهج الكمي، **مجلة العلوم التربوية والنفسيه_ البحرين**، المجلد السابع_ العدد الثاني.

Bauer, S., Lambert, M., & Nielsen, S. (2004). Clinical Significance Methods: A Comparison of Statistical Techniques. *Journal of Personality Assessment*, *82*(1), 60–70.

Campbell, T. C. (2005). An introduction to clinical significance: An alternative index of intervention effect for group experimental design. *Journal of Early Intervention, 27*, 210-227.

Cronbach, L., & Gleser, G. (1959). Interpretation of reliability and validity coefficients: remarks on a paper by Lord. *Journal of Educational Psychology, 50*, 230–237.

Gulliksen, H. (1950). *Theory of mental tests*. New York, NY: Wiley.

Hageman, W. J., & Arrindell, W. A. (1999). Establishing clinically significant change: increment of precision and the distinction between individual and group level analysis. *Behaviour Research and Therapy, 37*, 1169-1193.

Hansen, N., Lambert, M. J., & Forman, E. M. (2002). Comparisons of clinically significant change in clinical trials and naturalistic practice settings: The dose-effect relationship and its implication for practice. *Clinical Psychology: Science and Practice, 9*, 329–343.

Hsu, L. M. (1999). A comparison of three methods of identifying reliable and clinically significant client changes: commentary on Hageman and Arrindell. *Behaviour Research and Therapy, 37, 1195-1202.*

Hubbard, R., & Ryan, P. A. (2000). The historical growth of statistical significance testing in psychology-- and its future prospects. *Educational and Psychological Measurement, 60*, 661-681.

Huberty, C. J. (1999). On some history regarding statistical testing. In B. Thompson (Ed.), *Advances in social science methodology* (Vol. 5, pp. 1-23). Stamford, CT: JAI Press. Huberty, C. J. (2002). A history of effect size indices. *Educational and Psychological Measurement, 62, 227-240.*

Jacobson, N. S., Follette,W. C., & Revenstorf, D. (1984). Toward a standard definition of clinically significant change. *Behavior Therapy, 17,* 308–311.

Jacobson, N. S., & Truax, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology, 59,* 12–19.

Kendall, P. C. (1999). Clinical significance. *Journal of Consulting and Clinical Psychology, 67, 283-284.*

Kendall, P. C., Marrs-Garcia, A., Nath, S. R., & Sheldrick, R. C. (1999). Normative comparisons for the evaluation of clinical significance. *Journal of Consulting and Clinical Psychology, 67,* 285–299.

Lambert, M. J., Whipple, J. L., Smart, D.W., Vermeersch, D. A., Nielsen, S.L., & Hawkins, E. J. (2001). The effects of providing therapists with feedback on patient progress during psychotherapy: Are outcomes enhanced? *Psychotherapy Research, 11,* 49–68.

Lord, F., & Novick, M. (1968). *Statistical theories of mental scores. Reading*, MA: Addison-Wesley.

McGlinchey, J. B., & Jacobson, N. S. (1999). Clinically significant but impractical?: A response to Hageman and Arrindell. *Behavior Research and Therapy, 37, 1211-1217.*

Peterson, L. (2008). "*Clinical*" *Significance:* "*Clinical*" *Significance and* "*Practical*" *Significance are NOT the Same Things. Paper presented at the annual meeting of the Southwest Educational Research Association*, New Orleans, February 7.

Speer, D. C. (1992). Clinically significant change: Jacobson and Truax (1991) revisited. *Journal of Consulting and Clinical Psychology, 60*, 402-408.

Stevens, J.P. (1996). *Applied multivariate statistics for the social sciences* (3rd ed.). NJ: Lawrence Erlabaum.

Thompson, B. (1998). *Encouraging effect size reporting is not working: The etiology of research resistance to changing practices*. Paper presented at the annual meeting of the Southwest Educational Research Association (Houston, Tx, January, 1998). (Eric Document Reproduction Service No. ED 416214).

Thompson, B. (2006). *Foundations of behavioral statistics: An insight-based approach*. New York: The Guilford Press.

Thompson, B. (2002). "Statistical", "practical," and "clinical": How many kinds of significance do counselors need to consider? *Journal of Counseling and Development, 80, 64-80.*

Wilkinson, L., & APA Task Force on Statistical Inference. (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist, 54*, 594-604.